

**Title:** A systems approach identifies time-dependent associations of multimorbidities with pancreatic cancer risk

**Short title:** Multimorbidity and pancreatic cancer

**Authors:** P. Gomez-Rubio<sup>1</sup>, V. Rosato<sup>2,3</sup>, M. Marquez<sup>1</sup>, C. Bosetti<sup>4</sup>, E. Molina-Montes<sup>1</sup>, M. Rava<sup>1</sup>, J. Piñero<sup>5</sup>, C. W. Michalski<sup>6,7</sup>, A. Farré<sup>8</sup>, X. Molero<sup>9,10,11</sup>, M. Löhr<sup>12</sup>, L. Ilzarbe<sup>13</sup>, J. Perea<sup>14</sup>, W. Greenhalf<sup>15</sup>, M. O'Rourke<sup>16</sup>, A. Tardón<sup>17,11</sup>, T. Gress<sup>18</sup>, V. M. Barberà<sup>19</sup>, T. Crnogorac-Jurcevic<sup>20</sup>, L. Muñoz-Bellvís<sup>21</sup>, E. Domínguez-Muñoz<sup>22</sup>, A. Gutiérrez-Sacristán<sup>5</sup>, J. Balsells<sup>9,10,11</sup>, E. Costello<sup>15</sup>, C. Guillén-Ponce<sup>23</sup>, J. Huang<sup>12</sup>, M. Iglesias<sup>13</sup>, J. Kleeff<sup>6,15</sup>, B. Kong<sup>6</sup>, J. Mora<sup>8</sup>, L. Murray<sup>16</sup>, D. O'Driscoll<sup>24</sup>, P. Peláez<sup>14</sup>, I. Poves<sup>13</sup>, R. T. Lawlor<sup>26</sup>, A. Carrato<sup>23</sup>, , M. Hidalgo<sup>27</sup>, A. Scarpa<sup>26</sup>, L. Sharp<sup>24,25</sup>, L. I. Furlong<sup>5</sup>, F. X. Real<sup>28,29</sup>, C. LaVecchia<sup>2</sup>, N. Malats<sup>1</sup> on behalf of the PanGenEU Study Investigators<sup>30</sup>

#### Authors' affiliations

(1) Genetic and Molecular Epidemiology Group, Spanish National Cancer Research Center (CNIO), Madrid, and CIBERONC, Spain.

(2) Branch of Medical Statistics, Biometry and Epidemiology "G.A. Maccacaro", Department of Clinical Sciences and Community Health, University of Milan., Milan, Italy.

(3) Unity of Medical Statistics, Biometry and Bioinformatics, National Cancer Institute, IRCCS Foundation, Milan, Italy.

(4) Department of Epidemiology, Mario Negri Institute for Pharmacological Research-IRCCS, Milan, Italy.

(5) Research Programme on Biomedical Informatics (GRIB), Hospital del Mar Research Institute (IMIM), Pompeu Fabra University (UPF), Barcelona, Spain.

(6) Department of Surgery, Technical University of Munich, Munich, Germany.

(7) Department of Surgery, University of Heidelberg, Heidelberg, Germany.

(8) Department of Gastroenterology, Santa Creu i Sant Pau Hospital, Barcelona, Spain.

(9) Exocrine Pancreas Research Unit and Vall d'Hebron Research Institute (VHIR), Vall d'Hebron University Hospital, Barcelona, Spain.

- (10) Department of Medicine, Universitat Autònoma de Barcelona, Barcelona, Spain.
- (11) Centro de Investigación Biomédica en Red, Enfermedades Hepáticas y Digestivas, Epidemiológica y Salud Pública (CIBEREHD and CIBERESP), Madrid. Spain.
- (12) Gastrocentrum, Karolinska Institutet and University Hospital, Stockholm, Sweden.
- (13) Department of Gastroenterology, del Mar—Parc de Salut Mar University Hospital, Barcelona, Spain.
- (14) Department of Surgery, University Hospital 12 de Octubre, Madrid, Spain.
- (15) Department of Molecular and Clinical Cancer Medicine, The Royal Liverpool University Hospital, Liverpool, UK.
- (16) Centre for Public Health, Queen's University Belfast, Belfast, UK.
- (17) Department of Medicine, University Institute of Oncology of Asturias, Oviedo, Spain.
- (18) Department of Gastroenterology, University Hospital of Giessen and Marburg, Marburg, Germany.
- (19) Molecular Genetics Laboratory, General University Hospital of Elche , Elche, Spain.
- (20) Centre for Molecular Oncology, Barts Cancer Institute, Queen Mary University of London, John Vane Science Centre, London, UK.
- (21) General and Digestive Surgery Department, Salamanca University Hospital, Salamanca, Spain.
- (22) Department of Gastroenterology, University Clinical Hospital of Santiago de Compostela, Santiago de Compostela, Spain.
- (23) Department of Oncology, Ramón y Cajal Hospital, Madrid, and CIBERONC, Spain.
- (24) National Cancer Registry Ireland, Cork, Ireland,
- (25) Institute of Health & Society, Newcastle University, United Kingdom.
- (26) ARC-Net centre for Applied Research on Cancer and Department of Pathology and Diagnostics, University and Hospital trust of Verona, Verona, Italy.
- (27) Madrid-Norte-Sanchinarro Hospital, Madrid, Spain.
- (28) Epithelial Carcinogenesis Group, Spanish National Cancer Research Centre (CNIO), Madrid, and CIBERONC, Spain.
- (29) Department of experimental and health sciences, Pompeu Fabra University, Barcelona, Spain.
- (30) PanGenEU Study Investigators (Supplemental Annex S1).

65

66 **Abbreviations**

67 Pancreatic ductal adenocarcinoma, PDAC; Type 2 diabetes mellitus, T2DM; Metabolic syndrome,  
68 MetS; *Helocobacter pylori*, *H. pylori*; multimorbidity pattern, MP.

69

70 **Corresponding author**

71 Dr. Núria Malats,

72 Genetic and Molecular Epidemiology Group, Spanish National Cancer Research Center

73 (CNIO), C/Melchor Fernandez Almagro, 3, 28029, Madrid, Spain, Phone: +34-912-246-900

74 (ext.3330), Fax: +34-912-246-911, E-mail: nmalats@cnio.es

75

76

77 Abstract: 277 words

78 Text: 3680 words

79 References: 26

80 Tables: 2

81 Figures: 1

82 Supplementary material: Supplemental Annex, Supplementary Methods, 6 Supplementary

83 Tables, and 3 Supplementary Figure

## Abstract

## Background

Pancreatic ductal adenocarcinoma (PDAC) is usually diagnosed in late adulthood; therefore, many patients suffer or have suffered from other diseases. Identifying disease-patterns associated with PDAC risk may enable a better characterization of high-risk patients.

## Methods

Multimorbidity patterns (MPs) were assessed from 17 self-reported conditions using hierarchical clustering, principal component, and factor analyses in 1705 PDAC cases and 1084 controls from a European population. Their association with PDAC was evaluated using adjusted logistic regression models. Time since diagnosis of morbidities to PDAC diagnosis/recruitment was stratified into recent (<3 years) and long-term ( $\geq 3$  years). The MPs and PDAC genetic networks were explored with DisGeNET bioinformatics-tool which focuses on gene-diseases associations available in curated databases.

## Results

Three MPs were observed: gastric (heartburn, acid regurgitation, *H. pylori* infection, and ulcer), metabolic syndrome (obesity, type-2 diabetes, hypercholesterolemia, and hypertension), and atopic (nasal allergies, skin allergies, and asthma). Strong associations with PDAC were observed for  $\geq 2$  recently diagnosed gastric conditions (odds ratio [OR], 6.13; 95%CI 3.01-12.5) and for  $\geq 3$  recently diagnosed metabolic syndrome conditions (OR, 1.61; 95%CI 1.11-2.35). Atopic conditions were negatively associated with PDAC (high adherence score OR for tertile III, 0.45; 95%CI 0.36 – 0.55). Combining type-2 diabetes with gastric MP resulted in higher PDAC risk for recent (OR, 7.89; 95%CI 3.9-16.1) and long-term diagnosed conditions (OR, 1.86; 95%CI 1.29-2.67). A common genetic basis between MPs and PDAC was observed in the bioinformatics analysis.

## Conclusions

Specific multimorbidities aggregate and associate with PDAC in a time-dependent manner. A better characterization of a high-risk population for PDAC may help in the early diagnosis of this cancer. The common genetic basis between MP and PDAC points to a mechanistic link between these conditions.

## Keywords

Pancreatic cancer; Multimorbidity; Risk.

**Key Message (400characters max (spaces included)= 397)**

We explored the co-occurrence of morbidities and their common effect on pancreatic cancer. We showed that some groups of morbidities cluster together. Given that multimorbidity is frequent in the adult population, evaluating the collective effect of patterns of conditions will allow a better characterization of risk for this malignancy potentially improving the management of affected patients.

## Introduction

Pancreatic ductal adenocarcinoma (PDAC) is a dreadful disease usually diagnosed in late adulthood; hence, many patients suffer from other diseases [1,2]. PDAC patients present certain morbidities more often than non-PDAC patients, including chronic pancreatitis, type two diabetes mellitus (T2DM) or obesity [3–5]. Few studies focus in constellations of morbidities, and most explore associations between individual conditions and PDAC risk. In PDAC, multimorbidity has been mainly studied in the context of the metabolic syndrome (MetS), a cluster of conditions including abdominal obesity, hypertension, fasting hyperglycemia and serum hypertriglyceridemia, and low HDL cholesterol [6]. Having  $\geq 3$  MetS conditions has been associated with a 2-fold increased PDAC risk [7] yet, few studies have characterized in detail this relationship [8–12]. Additional multimorbidity patterns (MPs) affecting PDAC might exist, providing key information on disease mechanisms and contributing to the development of strategies to improve risk assessment.

We explored the patterns of co-occurrence of 17 self-reported medical conditions and their association with PDAC in a European case-control population. Identifying MPs and evaluating their effect on PDAC risk provides clues on common (genetic and/or environmental) backgrounds and allow us to obtain a finer characterization of risk. The assessment of time since diagnosis of these conditions allowed us to evaluate whether the morbidities represent a risk factor or an early manifestation of the malignancy.

## Methods

### *Study population*

The PanGenEU case-control study recruited PDAC patients  $\geq 18$  years old from Spain, United Kingdom, Germany, Ireland, Italy, and Sweden between 2009 and 2014 (Supplementary Annex S1, Table S2). Italy contributed with cases only. Controls were hospital-based, except in Ireland and Sweden where controls were population-based. Controls had no history of PDAC and principal diagnosis at admission was unrelated to known risk factors of PDAC (Supplementary Table S1). Response rate was 86.3% in cases and 77.8% in controls. IRB approval and written informed consent were obtained from all centers and participants.

Health monitors performed in-person interviews; subjects answered “yes/no” to “Has your doctor ever told you that you had any of the following illnesses, health problems or procedures?” for 25 medical conditions (Supplementary Table S3). Obesity (Body mass index, BMI  $\geq 30$  kg/m<sup>2</sup>)

was calculated using reported usual adult height and weight two years before recruitment. Only conditions with prevalence  $\geq 2\%$  were considered. Subjects without information in the medical questions were removed ( $n=268$ , 8.8%), leaving 1705 cases and 1084 controls. Age, sex and smoking were not statistically different between included and excluded subjects ( $p$ -value  $>0.05$ ).

### **Statistical analysis**

Data was imputed using missForest R package (Supplementary Methods). Adjustment variables were selected using the 10% change of estimate method. Statistical significance was considered as  $p$ -value  $<0.05$ . Analyses were performed using R version 3.1.2.

All combinations of two (comorbidity) and three morbidities (trimorbidity) within an individual were considered (Supplementary Methods). MPs were defined separately for cases and controls using hierarchical cluster analysis, principal component (PCA), and factor (FA) analyses (Supplementary Methods). Variables were created by combining presence/absence of conditions (0=none, 1=any one, 2=any two, 3=any three or more) for gastric (*H. pylori*, acid regurgitation, heartburn and ulcer), MetS (T2D, hypertension, obesity and hypercholesterolemia), and atopic (asthma, nasal and skin allergies) patterns. To determine the degree of a subject belonging to the identified patterns, i.e. adherence to a pattern, factor scores were calculated: low and high adherence for the first and third tertile, respectively. Logistic regressions were performed to evaluate the association of each MP and adherence tertiles with PDAC. Time since diagnosis of morbidities to PDAC diagnosis/recruitment was explored through stratification as recent ( $<3$  years) and long-term ( $\geq 3$  years). Variables whose associations resulted in an OR change  $>1.5$  after stratification by time since diagnosis were combined into a single variable for further analysis. Interaction between MPs and age, sex, and smoking were calculated comparing models with and without interaction terms through likelihood ratio test. Additive interactions of morbidities within each MP were also tested (Supplementary Methods). Excluding patients from Italy did not modify the results. Heterogeneity by country was not significant ( $p$ -value  $>0.2$ ). Internal validation of these models was also performed (Supplementary Methods)

Multimorbidity system analysis was performed with DisGeNET (Supplementary Methods).

## Results

Demographic and lifestyle information is provided in Supplementary Table S2.

### *Individual conditions.*

Number of morbidities reported was similar among cases and controls (~7% zero morbidities; ~38%  $\geq 4$  morbidities). When comparing subjects reporting zero morbidities with those reporting  $\geq 4$ , the former were younger (mean age 59 vs. 67 years), less heavy smokers (15.5% vs. 20.1%), and more commonly males (61.1% vs. 53.3%).

### *Comorbidity assessment.*

Most morbidities (81%) associated with each other in the same direction in cases and controls, with only moderate differences (Supplementary Figure 1). Nine comorbidities and four trimorbidities were significantly associated with PDAC after multiple test correction (Supplementary Table S4). The strongest positive association was observed for “T2DM and acid regurgitation” (OR, 4.25; 95%CI, 2.55-7.08). The average SPP for these associations was 93.3, i.e. if the analysis were repeated 100 times the association would be statistically significant 93 times after multiple test correction. Significant trimorbidities (average SPP, 93.4) always included T2DM and a combination of other conditions with ORs around 4, suggesting that the association of T2DM with PDAC risk might be potentiated by some morbidities.

### *Multimorbidity assessment.*

Three clear patterns of conditions emerged: 1) Gastric: heartburn, acid regurgitation, *H. pylori*, and ulcer; 2) MetS: obesity, T2DM, hypercholesterolemia and hypertension; and 3) Atopic: nasal allergies, skin allergies and asthma (Figure 1). The remaining conditions did not follow a consistent pattern of co-occurrence and were not considered in further analyses.

A positive association was observed between PDAC risk and the gastric and MetS patterns with a positive trend ( $p$ -value  $< 0.01$ ); a negative association was observed in the atopic pattern (Table 1). Similar findings were observed when comparing the highest with the lowest adherence tertiles. An interaction was observed between MetS and sex;  $\geq 3$  MetS conditions was associated with PDAC in males but not in females ( $p$ -value = 0.032, Supplementary Table S5).



219 *Time since diagnosis.*

220 Associations do not necessarily indicate causality; therefore, we assessed the  
221 associations in relation to the period since diagnosis of conditions and PDAC. Recent diagnosis  
222 of the gastric pattern showed stronger association with PDAC risk than long-term diagnosis (OR  
223 and 95%CI: 6.13, 3.01-12.5; 1.17, 0.79-1.73; respectively). Stratification of the MetS pattern was  
224 done based of T2DM diagnosis because the other conditions did not show different associations  
225 with PDAC by time since diagnosis: OR, 1.61; 95%CI, 1.11-2.35 for recent and OR, 1.35; 95%CI,  
226 0.99-1.85 for long-term diagnosis. This analysis could not be performed for the atopic pattern  
227 because time since diagnosis was only available for asthma.

228 Gastric MP and T2DM showed the greatest change in their association with PDAC after  
229 stratification by time since diagnosis (Supplementary Table S3). Considered altogether, having  
230  $\geq 1$  of these was strongly associated with PDAC among subjects with recent diagnosis (OR, 7.89;  
231 95%CI, 3.9-16.1 for  $\geq 2$  conditions) while a moderate estimate was observed among subjects with  
232 long-term diagnosis (OR, 1.86; 95%CI, 1.29-2.67 for  $\geq 3$  conditions, Table 2). The average AUC  
233 for this model was 0.78 (SD= 0.011) for recently diagnosed conditions; and 0.72 (SD= 0.01) and  
234 0.74 (SD= 0.009) for long-term and lifetime models, respectively (data not shown). Treating  
235 obesity as a confounder showed similar results (Supplementary Table S6).

236 Bioinformatics analysis showed a stronger genetic link among morbidities included in the  
237 metabolic and atopic patterns in comparison to those of the gastric pattern, possibly due to less  
238 reported disease-gene associations in the latter. All patterns shared genetic links with PDAC  
239 (Supplementary Figure S2), and many of these conditions shared inflammatory related genes  
240 such as *TNF*, *CXCL8*, *HIF1A*, and *PTGS2* (Supplementary Figure S3).

## 242 **Discussion**

243 We analyzed the aggregation and association of 17 self-reported medical conditions with PDAC  
244 risk in a large case-control European study. We identified three MPs: the gastric and MetS  
245 patterns, positively associated with PDAC, and the atopic pattern, negatively associated.  
246 Moreover, we explored MPs in the temporal context of pancreatic cancer which provides clues  
247 about common mechanisms and causal effect.

248           Among the gastric pattern, sensitivity analysis showed that heartburn and acid  
249 regurgitation had the strongest effect on the association. In the MetS pattern, T2DM was the main  
250 driving condition since removing T2DM from the pattern resulted in loss of significance. In the  
251 atopic pattern, none of the conditions seemed to drive the association individually.

252           Associations between gastric and MetS patterns with PDAC were significant only among  
253 subjects with recently diagnosed conditions. Our results strengthen previous findings with smaller  
254 case sample size or pooled heterogeneous populations reporting positive associations between  
255 history of gastric or duodenal ulcer with PDAC only among recently diagnosed subjects [2, 13,  
256 14]. Two cohort studies reported a significant association between PDAC and long-term diagnosis  
257 (up to 20 years) of gastric but not duodenal ulcer [15, 16]; whether gastric ulcer was related to *H.*  
258 *pylori* infection was not elucidated. A recent meta-analysis including ten case-control studies  
259 reported a non-significant association between overall *H. pylori* infection and PDAC [17]. These  
260 studies did not consider time since diagnosis and are limited in their ability to assess causality.  
261 Less information exists for acid regurgitation or heartburn. One study reported a significant  
262 positive association between PDAC and episodes of  $\geq 4$  weeks of heartburn up to 5 years prior to  
263 cancer diagnosis/interview [18]. For the MetS pattern, this observation was attributed to T2DM  
264 confirming and extending previous reports of a stronger positive association among new-onset  
265 diabetics [4, 19]. Importantly, the significant association between three or more MetS conditions  
266 and PDAC was restricted to males, pointing to a potential role of BMI since, contrary to females,  
267 there were more obese males among cases (22.6%) than controls (18.9%). Previous studies  
268 reported contradictory results on the interaction between MetS and sex [7, 8, 11, 12], more studies  
269 are needed to confirm our observations. Although recent asthma diagnosis was reported by <1%,  
270 long-term asthmatics were negatively associated with PDAC. Consistently, a significant inverse  
271 association between PDAC and nasal allergies unrelated to disease duration has been reported  
272 [20].

273           Five conditions (i.e. T2DM and gastric conditions) stood out by the magnitude of OR  
274 change after stratification by time since diagnosis. When consolidated into a single variable, a  
275 stronger positive association was observed between having  $\geq 1$  of these conditions recently  
276 diagnosed and PDAC. T2DM was an important driver of this association, but having  $\geq 3$  of these  
277 conditions showed a stronger estimate than by T2DM alone highlighting the potential significance  
278 of combining morbidities when studying PDAC. Bootstrapped AUCs showed a fair performance  
Gómez-Rubio et al. Multimorbidity and pancreatic cancer – Ms (March 19, 2016) 10

of the models including these five conditions; restriction to recent diagnosis resulted in the best performance. Conceivably, additive interactions between morbidities within MPs could explain these results. In this regard we observed a significant additive interaction between asthma and both nasal and skin allergies, and a significant sub-additive interaction between diabetes and the combination of any three or more gastric conditions; however, we were limited by sample size in these analyses. Mechanistically, these associations might be explained in the context of systemic inflammation: early events during PDAC carcinogenesis could favor the development of type 3c diabetes, altogether prompting dysbiosis which could, in turn, be manifested as gastric conditions. Further analysis showed an intricate genetic system among the MPs and pancreatic cancer pointing to a potential mechanistic link between these (Supplementary Figure S2 and S3).

It has been hypothesized that PDAC development might occur during the ten years previous to its detection with disease manifestations in the preceding 1-2 years. Yet, there are arguments in favor of a rapid "catastrophic" evolution shortly before diagnosis [21]. We aimed at discriminating medical conditions acting as true risk/protective factors (long-term diagnosis) vs. early manifestations of PDAC (recent diagnosis). We show that long-term T2DM and asthma are associated with PDAC, pointing to them as true risk and protective factors, respectively. Interestingly, the co-occurrence of long-term T2DM with long-term gastric conditions was associated with a stronger PDAC risk than that conferred by T2DM alone indicating that while T2DM is a main risk factor of PDAC, its overlap with other gastric conditions might further aggravate carcinogenesis. We also provide strong evidence that several conditions associate with PDAC risk only when diagnosed shortly before PDAC, suggesting that these could be PDAC manifestations. This concept has been discussed in the context of T2DM [4, 13, 19, 22, 23], and it is further supported by studies showing improvement of insulin resistance and glucose intolerance after PDAC resection [24, 25]. However, we cannot rule out that pathophysiologically distinct subtypes of T2DM and/or other conditions might accelerate PDAC progression resulting in a shorter time between diagnoses.

Early PDAC diagnosis provides the only opportunity for long-term survival. Establishing a set of conditions alerting the clinical staff of potential PDAC cases could improve diagnosis, management and/or treatment of patients and consequently, the outcome of the disease. However, the symptoms/conditions here described are relatively non-specific hindering their immediate adscription to PDAC. Moreover, whether PDAC stage at diagnosis modifies the

reported associations deserves further investigation. While innovative and detailed analyses have been performed to both identify MPs and to assess their risk with PDAC, bias by unmeasured confounders is always a concern (i.e., metformin or statins medications have been associated with decreased risk of PDAC). Future studies should focus on exploring the role of particular treatments in these associations.

Our study has some limitations. The information was self-reported thus, misclassification and bias cannot be completely ruled out; nevertheless, the consistency of basic findings in this report with existing literature argues against this possibility. Our study is one of the largest performed but - despite its size and the restriction of our analyses to conditions with a frequency  $\geq 2\%$  in cases and controls – statistical power remains an issue when considering multimorbidities and stratification by time since diagnosis.

To our knowledge, this is the first study that simultaneously considers the association of many morbidities with PDAC. We report three main MPs significantly associated with PDAC and show that the associations change depending on the time since diagnosis of morbidities. Owing the high prevalence of multimorbidities, evaluating MPs may help to improve the characterization of PDAC risk. Confirmation of these results in independent studies will be critical in their generalization and future clinical application. Identification of early manifestations of PDAC may help define subpopulations of patients in whom screening might be cost-effective [26]. Moreover, the discovery of risk and protective associations should lead to more pathophysiological studies. Thus, the knowledge generated could contribute to implement novel preventive and therapeutic strategies.

## Acknowledgments

We acknowledge the COST Action BM1204 EUPancreas "An integrated European platform for pancreatic cancer research: from basic science to clinical and public health interventions for a rare disease", coordinators, field and administrative workers, technicians and study participants of the *European Study into Digestive Illnesses and Genetics* (PanGenEU) study.

## **Funding**

The work was partially supported by Fondo de Investigaciones Sanitarias (FIS), Instituto de Salud Carlos III-FEDER, Spain (#PI11/01542, #PI0902102, #PI12/01635, #PI12/00815, #PI13/00082, CP10/00524); Red Temática de Investigación Cooperativa en Cáncer, Spain (#RD12/0036/0034, #RD12/0036/0050, #RD12/0036/0073); World Cancer Research (WCR #15-0391); Acción Especial de Genómica, Spain (#GEN2001-4748-c05-03); *EU-6FP Integrated Project* (#018771-MOLDIAG-PACA), EU-FP7-HEALTH (#259737-CANCERALIA, #256974-EPC-TM-Net, #602783-Cam-Pac), Associazione Italiana Ricerca sul Cancro (AIRC #12182); Cancer Focus Northern Ireland and Department for Employment and Learning; ALF (#SLL20130022), Sweden; Italian Foundation for Cancer Research (FIRC). The funders had no role in the study design or the analysis and interpretation of the data.

## **Disclosure**

The authors have declared no conflicts of interest.

## References

1. Olson SH. Selected medical conditions and risk of pancreatic cancer. *Mol. Carcinog.* 2012; 51(1):75–97.
2. Ko AH, Wang F, Holly E a. Pancreatic cancer and medical history in a population-based case-control study in the San Francisco Bay Area, California. *Cancer Causes Control* 2007; 18(8):809–819.
3. Raimondi S, Lowenfels AB, Morselli-Labate AM et al. Pancreatic cancer in chronic pancreatitis; aetiology, incidence, and early detection. *Best Pract. Res. Clin. Gastroenterol.* 2010; 24(3):349–358.
4. Bosetti C, Rosato V, Li D et al. Diabetes, antidiabetic medications, and pancreatic cancer risk: an analysis from the International Pancreatic Cancer Case-Control Consortium. *Ann. Oncol.* 2014; 25(10):2065–2072.
5. Genkinger JM, Spiegelman D, Anderson KE et al. A pooled analysis of 14 cohort studies of anthropometric factors and pancreatic cancer risk. *Int. J. Cancer* 2011; 129(7):1708–1717.
6. Eckel RH, Alberti K, Grundy SM, Zimmet PZ. The metabolic syndrome. *Lancet* 2010; 375(9710):181–183.
7. Rosato V, Tavani A, Bosetti C et al. Metabolic syndrome and pancreatic cancer risk: A case-control study in Italy and meta-analysis. *Metabolism.* 2011; 60(10):1372–1378.
8. Russo A, Autelitano M, Bisanti L. Metabolic syndrome and cancer risk. *Eur. J. Cancer* 2008; 44(2):293–297.
9. Inoue M, Noda M, Kurahashi N et al. Impact of metabolic factors on subsequent cancer risk: results from a large-scale population-based cohort study in Japan. *Eur. J. Cancer Prev.* 2009; 18(3):240–247.
10. Matthews CE, Sui X, Lamonte MJ et al. Metabolic syndrome and risk of death from cancers of the digestive system. *Metabolism* 2010; 59(8):1231–1239.
11. Johansen D, Stocks T, Jonsson H et al. Metabolic factors and the risk of pancreatic cancer: a prospective analysis of almost 580,000 men and women in the Metabolic Syndrome and Cancer Project. *Cancer Epidemiol. Biomarkers Prev.* 2010; 19(9):2307–2317.
12. Stocks T, Bjorge T, Ulmer H et al. Metabolic risk score and cancer risk: pooled analysis of seven cohorts. *Int. J. Epidemiol.* 2015:1–11.
13. Lipworth L, Zucchetto A, Bosetti C et al. Diabetes mellitus, other medical conditions and pancreatic cancer: a case-control study. *Diabetes Metab Res Rev* 2011; 27(3):255–261.
14. Bosetti C, Lucenteforte E, Bracci PM et al. Ulcer, gastric surgery and pancreatic cancer risk: an analysis from the International Pancreatic Cancer Case-Control Consortium (PanC4). *Ann. Oncol.* 2013; 24(11):2903–2910.
15. Luo J, Nordenvall C, Nyrén O et al. The risk of pancreatic cancer in patients with gastric or duodenal ulcer disease. *Int. J. Cancer* 2007; 120(2):368–372.
16. Bao Y, Spiefelman D, Li R et al. History of peptic ulcer diseases and pancreatic cancer risk in men. *Gastroenterology* 2010; 138(2):541–9.
17. Schulte A, Pandeya N, Fawcett J et al. Association between *Helicobacter pylori* and pancreatic cancer risk: a meta-analysis. *Cancer Causes Control* 2015; 26(7):1027–1035.
18. Holly EA, Chaliha I, Bracci PM, Gautam M. Signs and symptoms of pancreatic cancer: a population-based case-control study in the San Francisco Bay area. *Clin. Gastroenterol. Hepatol.* 2004; 2(6):510–517.

402 19. Batabyal P, Vander Hoorn S, Christophi C, Nikfarjam M. Association of diabetes mellitus  
403 and pancreatic adenocarcinoma: a meta-analysis of 88 studies. *Ann. Surg. Oncol.* 2014;  
404 21(7):2453–62.

405 20. Cotterchio M, Lowcock E, Hudson TJ et al. Association between allergies and risk of  
406 pancreatic cancer. *Cancer Epidemiol. Biomarkers Prev.* 2014; 23(3):469–80.

407 21. Real FX. A “catastrophic hypothesis” for pancreas cancer progression. *Gastroenterology*  
408 2003; 124(7):1958–1964.

409 22. Ben Q, Xu M, Ning X et al. Diabetes mellitus and risk of pancreatic cancer: A meta-  
410 analysis of cohort studies. *Eur. J. Cancer* 2011; 47(13):1928–1937.

411 23. Elena JW, Steplowski E, Yu K et al. Diabetes and risk of pancreatic cancer: a pooled  
412 analysis from the pancreatic cancer cohort consortium. *Cancer Causes Control* 2013;  
413 24(1):13–25.

414 24. Pannala R, Leirness JB, Bamlet WR et al. Prevalence and Clinical Profile of Pancreatic  
415 Cancer-Associated Diabetes Mellitus. *Gastroenterology* 2008; 134(4):981–987.

416 25. Nakamori S, Ishikawa O, Ohigashi H et al. Increased blood proinsulin and decreased C-  
417 peptide levels in patients with pancreatic cancer. *Hepatogastroenterology* 1999;  
418 46(25):16–24.

419 26. Gerdtsson AS, Malats N, Säll A et al. A Multicenter Trial Defining a Serum Protein  
420 Signature Associated with Pancreatic Ductal Adenocarcinoma. *Int. J. Proteomics* 2015;  
421 2015:587250.

422

423

424

425  
426

**Supplementary Annex.** PanGenEU centres and investigators (<sup>1</sup> Principal Investigator in each centre)

Spanish National Cancer Research Centre (CNIO), Madrid, Spain: Núria Malats<sup>1</sup>, Francisco X Real<sup>1</sup>, Mirari Marquez, Roger Milne, Paulina Gómez, Ana Sagrera.

Verona University, Italy: Aldo Scarpa<sup>1</sup>, Rita Lawlor, Giada Bonizzato, Sonia Grimaldi.

National Cancer Registry Ireland, Cork, Ireland: Linda Sharp<sup>1</sup>, Damian O'Driscoll.

Hospital Madrid-Norte-Sanchinarro, Madrid, Spain: Manuel Hidalgo<sup>1</sup>.

Hospital Ramon y Cajal, Madrid, Spain: Alfredo Carrato<sup>1</sup>, Carmen Guillén-Ponce.

Hospital del Mar, Barcelona, Spain: Lucas Ilzarbe<sup>1</sup>, Cristina Álvarez-Urturi, Xavier Bessa, Felipe Bory, Lucía Márquez Mosquera, Ignasi Poves Prim, Fernando Burdío, Luis Grande, Mar Iglesias, Javier Gimeno.

Hospital Vall d'Hebron, Barcelona, Spain: Xavier Molero<sup>1</sup>, Luisa Guarner, Joaquin Balcells.

Technical University of Munich, Germany: Christoph Michalski<sup>1</sup>, Jörg Kleeff, Bo Kong

Karolinska Institute, Stockholm, Sweden: Matthias Löhr<sup>1</sup>, Jiaqi Huang.

Hospital 12 de Octubre, Madrid, Spain: José Perea<sup>1</sup>, Pablo Peláez.

Hospital de la Santa Creu i Sant Pau, Barcelona, Spain: Antoni Farré<sup>1</sup>, Josefina Mora, Marta Martín, Vicenç Artigas, Carlos Guarner Argente, Francesc J Sancho, Mar Concepción, Teresa Ramón y Cajal.

The Royal Liverpool University Hospital, UK: William Greenhalf<sup>1</sup>, Eithne Costello.

Queen's University Belfast, UK: Liam Murray<sup>1</sup>, Michael O'Rorke, Marie Cantwell.

Laboratorio de Genética Molecular, Hospital General Universitario de Elche, Spain: Víctor M Barberá<sup>1</sup>, Javier Gallego.

Instituto Universitario de Oncología del Principado de Asturias, Oviedo, Spain: Adonina Tardón<sup>1</sup>, Luis Barneo.

Hospital Clínico Universitario de Santiago de Compostela, Spain: Enrique Domínguez Muñoz<sup>1</sup>, Antonio Lozano, Maria Luaces.

Hospital Clínico Universitario de Salamanca, Spain: Luís Muñoz-Bellvís<sup>1</sup>, J.M. Sayagués Manzano, M.L. Gutiérrez Troncoso, A. Orfao de Matos.

University of Marburg, Department of Gastroenterology, Phillips University of Marburg, Germany: Thomas Gress<sup>1</sup>, Malte Buchholz, Albrecht Neesse.

Queen Mary University of London, UK: Tatjana Crnogorac-Jurcevic<sup>1</sup>, Hemant M Kocher, Satyajit Bhattacharya, Ajit T Abraham, Darren Ennis, Thomas Dowe, Tomasz Radon

Scientific advisors of the PanGenEU Study: Debra T Silverman (NCI, USA) and Douglas Easton (U. of Cambridge, UK)

427  
428  
429



## Figure Legends

**Figure 1. Clustering of morbidities in pancreatic ductal adenocarcinoma cases and controls.** A) Dendrogram depicting Hierarchical cluster analysis using Yule's Q distance matrix, B) Loadings from principal component analysis, C) Loadings from factor analysis. Loadings  $\geq 0.3$  are considered to belong to the same component and factor, respectively. Yule's Q distance was calculated as  $(1 - ((ad-bc)/(ad+bc)))$ , where a indicates presence of both conditions, b and d absence of one condition and presence of the other, and c absence of both conditions. PanGenEU, 2009-2014.